REMARKS

Introduction

Claims 1-19 are pending. Claims 3, 12 and 15 have been amended to provide proper antecedent basis. Claims 14 and 15 have been amended to correct a typographical error. Claims 1, 6, 10, 13, and 16 have been amended to explicitly recite what was implicitly recited by these claims. For the reasons discussed in detail below, all of the pending claims are in condition for allowance.

Obviousness Rejections

The Office action has rejected all the claims under 35 U.S.C. § 103(a) as being obvious. The following table lists the claims and the relied-upon references.

Claims	References
1, 7, 10, and 17	Hearst, Cortes
6, 13, 16	Hearst
2-3, 11-12, and 14-15	Hearst, Cortes
4-5, 8-9, and 18-19	Hearst, Cortes

Applicants respectfully traverse these rejections. In the following, applicants provide an overview of their invention and of the primary relied-upon references and then discuss their differences.

Applicants' technique for modeling a data set uses a variation of a probabilistic model known as a Relevance Vector Machine (RVM). This variation uses product approximations, including the distribution of hyperparameters of the RVM, to obtain a posterior distribution. The RVM uses a prior distribution for the data set to infer the resulting prediction model. This prior distribution is determined from selection of an

initial set of hyperparameters. In this version of Applicants' technique, the RVM generates a separate distribution for each hyperparameter and iteratively updates the distribution of the set of hyperparameters, the distribution of the set of weights, and the distribution of the set of predetermined additional parameters. The iteration ceases upon reaching a chosen convergence criterion. Applicants' technique results in constructing a model that outputs a posterior distribution over both parameters and hyperparameters that is used for probabilistic prediction of an event(s) or expected behavior(s) for a given input(s).

Hearst presents an overview of a Support Vector Machine (SVM) and describes the optimal hyperplane algorithm used for SVM classification. Specifically, Hearst describes SVM classifiers that are based on a class of hyperplanes corresponding to decision functions and defines the optimal hyperplane as one that yields the maximal margin of separation between two classes. Such optimal hyperplanes are constructed using a subset of support vectors and are solved using constrained quadratic optimization. Hearst presents a geometrical illustration of a simple SV classifier for separating two toys, balls from diamonds. In this illustration, Hearst explains how the SVM maps training data nonlinearly into a high-dimensional feature space by constructing a separating hyperplane with maximal margin. Hearst describes this optimal hyperplane as orthogonal to the shortest line connecting the convex hulls of the two classes and intersects it half way. The margin of separation is maximized by minimizing the distance vector between nearest points of the convex hulls and the hyperplane. This results in a nonlinear decision boundary to separate examples from the two classes.

Cortes teaches increasing the capacity of a learning machine until the asymptotic error rate between the training error and the test error is within acceptable limits of accuracy. The capacity of the learning machine is increased by increasing the free parameters to more completely model the training data set. A learning machine is first trained using a training data set, and the training error is calculated. The trained learning machine is then tested using a test data set, and the test error is calculated. The asymptotic error rate is then calculated as the mean of the training error and the test error. Until the asymptotic error rate between the training error and the test error is within acceptable limits of accuracy, the capacity of the learning machine is increased, trained and tested iteratively.

The SVM in Hearst and the learning machines in Cortes provides predictions of an event or expected behavior that is not probabilistic. SVM expresses predictions in terms of a linear combination of kernel functions centered on a subset of the training data, known as support vectors. SVMs make explicit classifications using point predictions for new inputs and do not generate predictive distributions. Applicants' variational RVM is a probabilistic model of a learning machine that outputs a posterior distribution for use in probabilistic prediction.

The following table lists parts of claim 1 and the corresponding sections of the prior art upon which the Office action relies to allege obviousness.

	Claim 1	Prior Art
A.	selecting an initial set of hyperparameters for determining a prior distribution for the data set for modeling thereof, the prior distribution approximated by a product of a distribution of the set of	Hearst, Pg 18, right column for Hyperplane classifiers for hyperparameters; Hearst, pg 19 Figures 1 and 2 for hyperplane and hyperparameters and weights given to the distribution;

	hyperparameters, a distribution of a set of weights, and a distribution of a set of predetermined additional parameters	Hearst, pg 22, right column for distribution of weights for learning vector machines; Hearst, pg 22, Table I and left column for additional parameters for different learning algorithms.
B.	interactively updating the distribution of the set of weights, the distribution of the set of hyperparameters, and the distribution of the set of predetermined additional parameters until a predetermined convergence criterion has been reached	Hearst, pg 19, left column for training patterns that iteratively train the learning machine by updating the weights and parameters; Hearst, pg 19, left column for the final decision function or the convergence criterion to end the iterations.
C.	such that the product of the distribution of the set of hyperparameters, the distribution of the set of weights, and the distribution of the set of predetermined additional parameters as have been iteratively updated approximates the posterior distribution for modeling of the data set for probabilistic prediction	Hearst, pg 19 Figure 1 for the weight vector; Hearst, pg 18, left column for distribution functions and statistical analysis, the posterior distribution is a type of probability distribution and statistical function.

Applicants respectfully submit that claims 1-5, and 10-15 by similar analysis, are not rendered obvious by the relied-upon references, and that the Office action has incorrectly interpreted the relied-upon portions.

First, part A of claim 1 recites "selecting an initial set of hyperparameters for determining a prior distribution for the data set for modeling thereof, the prior distribution approximated by a product of a distribution of the set of hyperparameters, a distribution of a set of weights, and a distribution of a set of predetermined additional parameters." The relied-upon portions of Hearst relate to SVM classifiers that are based on a class of hyperplanes corresponding to decision functions. Neither a prior distribution nor an

initial set of hyperparameters for determining a prior distribution are identified in Hearst. The hyperplane classifiers according to the Office action's incorrect interpretation are the hyperparameters used for approximating a prior distribution of the data set. However, the hyperplane classifiers described by Hearst are constructed using a subset of support vectors of an SVM and are not used for approximating a prior distribution.

Second, part B of claim 1 recites "interactively updating the distribution of the set of weights, the distribution of the set of hyperparameters, and the distribution of the set of predetermined additional parameters until a predetermined convergence criterion has been reached." However, the portion of Hearst relied-upon in the Office action relates to constructing optimal hyperplanes by solving a constrained quadratic optimization problem using a subset of support vectors with associated weights. There is no identified distribution of a set of hyperparameters. There is no identified iteration process updating the distribution of the set of hyperparameters. The training patterns, to which the Office action refers, are described by Hearst as support vectors which are used to maximize the margin of separation between classes by minimizing the distance between nearest points of the convex hulls of the classes. This is accomplished by rescaling the weight vector and threshold parameters associated with the support vectors. Hearst's construction of optimal hyperplanes does not involve "interactively updating the distribution of the set of weights, the distribution of the set of hyperparameters" as recited by claim 1.

Third, part C of claim 1 recites "such that the product of the distribution of the set of hyperparameters, the distribution of the set of weights, and the distribution of the set of predetermined additional parameters as have been iteratively updated approximates

the posterior distribution for modeling of the data set for probabilistic prediction." Applicants can find nothing in Hearst or Cortes that describes a product of distributions that approximates a posterior distribution for modeling a data set for probabilistic prediction.

The following table lists parts of claim 6 and the corresponding sections of the prior art upon which the Office action relies to allege obviousness.

	Claim 6	Prior Art
A.	inputting a data set to be modeled	Inputting data is inherent to modeling a data set.
B.	determining a relevance vector learning machine via a variational approach to obtain a posterior distribution for the data set for probabilistic prediction	Hearst, pg 18, left column for determining distribution functions and statistical analysis in a vector learning machine, the posterior distribution is a type of probability distribution and statistical function.
C.	outputting at least the posterior distribution for the data set	Outputting calculated values is inherent to modeling data.

Applicants respectfully submit that claims 6-9, and 16-19 by similar analysis, are not rendered obvious by the relied-upon references, and that the Office action has incorrectly interpreted the relied-upon portions.

Part B of claim 6 recites "determining a relevance vector learning machine via a variational approach to obtain a posterior distribution for the data set for probabilistic prediction." As previously mentioned, the relied-upon portions of Hearst relate to SVM classifiers that are based on a class of hyperplanes corresponding to decision functions. Neither a relevance vector learning machine nor a posterior distribution are identified. Applicants can find nothing in Hearst or Cortes that describes a relevance vector

learning machine that outputs a posterior distribution for the data set for probabilistic prediction.

To establish *prima facie* obviousness of a claimed invention, all of the claim limitations must be taught or suggested by the prior art; (*In re Royka*, 490 F.2d 981, 180 USPQ 580 (CCPA 1974)), and "all words in a claim must be considered in judging the patentability of that claim against the prior art;" (*In re Wilson*, 424 F.2d 1382, 1385, 165 USPQ 494, 496 (CCPA 1970)). For at least the foregoing reasons, factual and legal, applicants submit that neither Hearst nor Cortes, whether considered alone or in any permissible combination, meet this requirements, and thus that the present Office action has failed to establish *prima facie* obviousness as a matter of law with respect to any of the claimed subject matter. Reconsideration and withdrawal of the rejections of pending claims based on Hearst and/or Cortes is respectfully requested.

Non-statutory Subject Matter Rejections

The Office action has also rejected claims 1-12 and 14-19 under 35 U.S.C. § 101 as being directed to non-statutory subject matter. Applicants strongly disagree with this rejection.

Applicants' technique constructs a probabilistic model that takes real-world data and outputs a posterior distribution used for probabilistic prediction of an event or expected behavior. Applicants have also amended claims 1, 6, 10, 13, and 16 to make explicit what was implicitly recited by these claims. In particular, these claims recited "for modeling of the data set," "outputting the posterior distribution for the data set" or similar language. These claims now explicitly recite "for probabilistic prediction." Thus,

the claimed invention as a whole clearly accomplishes a practical application, as it produces a useful, concrete and tangible result, and does so without pre-empting other uses of the mathematical principle behind it. *State Street Bank & Trust Co. v. Signature Financial Group Inc.*, 149 F. 3d 1368, 1374, 47 USPQ2d 1596, 1601-02 (Fed. Cir. 1998); *AT&T Corp. v. Excel Communications, Inc.*, 172 F.3d 1352, 1358, 50 USPQ2d 1447, 1452(Fed. Cir. 1999). Reconsideration and withdrawal of the 35 U.S.C. §101rejections is respectfully requested.

Conclusion

Based upon the above amendments and remarks, applicants respectfully request reconsideration and withdrawal of the rejections and timely allowance of this application.

Respectfully submitted,

Albert S. Michalik, Registration No. 37,395
Attorney for Applicants
Law Offices of Albert S. Michalik, PLLC
704 - 228th Avenue NE
Suite 193
Sammamish, WA 98074
(425) 836-3030 (telephone)
(425) 836-8957 (facsimile)

Received from < 425 836 8957 > at 6/14/03 7:11:57 PM [Eastern Daylight Time]